USN

Internal Assessment Test 2 – November 2024

Sub:	**Introduction to AI & ML**					Sub Code:	**21CS752**	Branch:	**ECE**
Date:	**19/11/24**	Duration:	**90 minutes**	Max Marks:	**50**	Sem/Sec:	**VII**		**OBE**

		Answer any FIVE FULL Questions	**MARKS**	**CO**	**RBT**
1		Describe the terms Artificial Intelligence (AI), Machine Learning (ML) and Deep Learning.	[10]	02	L1
2	a	Differentiate between Data Mining and Machine Learning	[5]	02	L2
	b	Illustrate the different types of Learning/ Training models in ML.	[5]	02	L1
3	a	Define any four techniques of data visualization.	[6]	02	L1
	b	Elucidate any two dimensionality reduction techniques.	[4]	03	L1
4	a	Let the data points be $\begin{pmatrix}2\\6\end{pmatrix}$ and $\begin{pmatrix}3\\7\end{pmatrix}$. Apply PCA and find the transformed data.	[10]	03	L3
5	a	What do you mean by hypothesis testing ? Explain giving examples.	[10]	03	L1
6	a	Define t-test giving an example.	[4]	03	L1
	b	Distinguish between supervised and unsupervised machine learning?	[6]	02	L2

Faculty CCI HOD

USN

Internal Assessment Test 2 – November 2024

Sub:	**Introduction to AI & ML**					Sub Code:	**21CS752**	Branch:	**ECE**
Date:	**19/11/24**	Duration:	**90 minutes**	Max Marks:	**50**	Sem/Sec:	**VII**		**OBE**

		Answer any FIVE FULL Questions	**MARKS**	**CO**	**RBT**
1		Describe the terms Artificial Intelligence (AI), Machine Learning (ML) and Deep Learning.	[10]	02	L1
2	a	Differentiate between Data Mining and Machine Learning	[5]	02	L2
	b	Illustrate the different types of Learning/ Training models in ML.	[5]	02	L1
3	a	Define any four techniques of data visualization.	[6]	02	L1
	b	Elucidate any two dimensionality reduction techniques.	[4]	03	L1
4	a	Let the data points be $\begin{pmatrix}2\\6\end{pmatrix}$ and $\begin{pmatrix}3\\7\end{pmatrix}$. Apply PCA and find the transformed data.	[10]	03	L3
5	a	What do you mean by hypothesis testing ? Explain giving examples.	[10]	03	L1
6	a	Define t-test giving an example.	[4]	03	L1
	b	Distinguish between supervised and unsupervised machine learning?	[6]	02	L2

Faculty CCI HOD

Solution

Ans-1 **Machine Learning (ML)**

Definition: Machine Learning is a subset of AI that focuses on developing algorithms and statistical models that enable machines to learn from data and improve their performance over time without being explicitly programmed.

Key Characteristics:

- Relies on data-driven approaches to make predictions or decisions.
- Types of ML:
    - **Supervised Learning**: Learns from labeled data.
    - **Unsupervised Learning**: Discovers patterns in unlabeled data.
    - **Reinforcement Learning**: Learns by interacting with the environment and receiving feedback.

Examples:

- Spam email filtering.
- Predictive analytics (e.g., forecasting sales).
- Recommender systems (e.g., Netflix recommendations).

Deep Learning (DL)

Definition: Deep Learning is a specialized subset of Machine Learning that uses multi-layered artificial neural networks to model and solve complex problems. These networks attempt to mimic the way the human brain processes information.

Key Characteristics:

- Can automatically extract features from raw data, reducing the need for manual feature engineering.
- Requires large datasets and significant computational resources.
- Performs exceptionally well in tasks like image recognition, speech recognition, and natural language processing.

Examples:

- Image recognition (e.g., Google Lens).
- Language translation (e.g., Google Translate).
- Autonomous vehicles (e.g., Tesla's Autopilot).

Relationship Among AI, ML, and DL

1. **Artificial Intelligence (AI)** is the broadest concept, encompassing any system that exhibits intelligent behavior.
2. **Machine Learning (ML)** is a subset of AI that focuses on learning from data to improve performance over time.
3. **Deep Learning (DL)** is a subset of ML, specifically dealing with neural networks that learn and make decisions from large amounts of unstructured data.

Ans-2 (a) While **Data Mining** and **Machine Learning** both deal with extracting valuable insights from data, they differ in terms of their goals, processes, and methods. Below is a detailed comparison:

Aspect	**Data Mining**	**Machine Learning**
Definition	Data Mining refers to the process of discovering patterns, relationships, and trends in large datasets using statistical, mathematical, and computational techniques.	Machine Learning is a subset of AI focused on developing algorithms that allow systems to learn from data and make predictions or decisions based on it.
Goal	To identify hidden patterns, trends, and associations in large datasets.	To build models that can generalize from data and make predictions or decisions without being explicitly programmed.
Approach	Primarily focused on **exploratory data analysis**, uncovering patterns or anomalies in data.	Focuses on **training algorithms** to make predictions or decisions based on past data.
Techniques	Techniques include **association rule mining**, **clustering**, **classification**, and **anomaly**	Techniques include **supervised learning** (e.g., regression, classification), **unsupervised**

Aspect	Data Mining	Machine Learning
	detection.	**learning** (e.g., clustering), and **reinforcement learning**.
Data Processing	Data mining often involves pre-processing and cleaning data before extracting meaningful patterns.	Machine learning algorithms also require preprocessing, but the focus is on using data to train models for predictions.
Result/Output	The output of data mining is usually **patterns**, **rules**, or **clusters**.	The output of machine learning is typically a **predictive model** or **classifier** that can be used for future predictions.
Focus	Primarily concerned with **exploring** and **summarizing** historical data.	Focuses on **modeling** and **learning** from data to improve predictions over time.
Data Dependency	Data mining can work on large datasets with mixed types of data (structured or unstructured).	Machine learning requires large, often labeled datasets (especially for supervised learning) to train models.
Nature of Learning	Data mining is generally **pattern discovery** in data without the need for explicit feedback or learning over time.	Machine learning is about **improving models** over time by learning from data through feedback (in supervised or unsupervised contexts).
Example Use Cases	- Market Basket Analysis (finding associations between products). - Customer segmentation (clustering).	- Spam email filtering (classification). - Predicting stock prices (regression). - Speech recognition (supervised learning).

Ans- 2(b) Types of Learning/Training Models in Machine Learning

Machine Learning (ML) is categorized into various types based on how the model learns from data. The primary types of learning models are **Supervised Learning**, **Unsupervised Learning**, **Semi-supervised Learning**, **Reinforcement Learning**, and **Self-supervised Learning**. Below is an explanation of each:

---

1. Supervised Learning

Definition: In supervised learning, the model learns from labeled data, where the correct output (target) is provided for each input. The goal is to learn a mapping from inputs to outputs so the model can make predictions for new, unseen data.

How it works:

- The model is trained on a **training dataset** that includes both the input features and their corresponding labels (targets).
- After training, the model can make predictions on new data and compare them with the true outputs to measure performance.

Types:

- **Regression**: Predicting continuous values (e.g., predicting house prices).
- **Classification**: Predicting discrete categories (e.g., identifying if an email is spam or not).

Examples:

- **Spam Detection** (classification): Identifying whether an email is spam or not based on labeled data of past emails.
- **House Price Prediction** (regression): Predicting the price of a house based on features like size, location, and age of the house.

---

2. Unsupervised Learning

Definition: In unsupervised learning, the model is given data without explicit labels (unlabeled data). The goal is to find patterns, structures, or relationships in the data without knowing the correct output beforehand.

How it works:

- The model analyzes the input data and groups or clusters similar data points together or finds hidden structures within the data.

- No supervision is provided, meaning there is no target output to compare the results against.

Types:

- **Clustering**: Grouping similar data points (e.g., customer segmentation).
- **Dimensionality Reduction**: Reducing the number of features in a dataset while retaining important information (e.g., Principal Component Analysis or PCA).

Examples:

- **Customer Segmentation** (clustering): Grouping customers based on purchasing behavior to target marketing campaigns.
- **Anomaly Detection** (clustering): Identifying unusual behavior in a dataset, like fraud detection in transactions.

---

3. Semi-supervised Learning

Definition: Semi-supervised learning lies between supervised and unsupervised learning. In this approach, the model is provided with a small amount of labeled data and a large amount of unlabeled data. The goal is to leverage the labeled data to guide learning and infer the relationships in the unlabeled data.

How it works:

- The model starts with the small labeled dataset to learn the underlying structure and generalize to the unlabeled data, gradually improving the model's accuracy by making use of both labeled and unlabeled data.

Examples:

- **Image Labeling**: Using a few labeled images (e.g., labeled as "cat" or "dog") and a larger set of unlabeled images to build an image classifier.
- **Text Classification**: Classifying documents with only a few labeled examples and many unlabeled texts.

---

4. Reinforcement Learning

Definition: Reinforcement learning (RL) is a type of learning where an agent learns how to make decisions by performing actions in an environment and receiving feedback in the form of rewards or penalties. The agent's goal is to maximize its cumulative reward over time.

How it works:

- The agent interacts with the environment, takes actions, and receives feedback (rewards or penalties).
- The agent then adjusts its actions based on past experiences to maximize its long-term cumulative reward.
- **Exploration vs. Exploitation**: The agent balances exploring new actions and exploiting known ones that have provided the most reward.

Types:

- **Model-based RL**: The agent builds a model of the environment and uses it for planning future actions.
- **Model-free RL**: The agent learns directly from interactions without explicitly modeling the environment.

Examples:

- **Game Playing** (e.g., AlphaGo): An AI agent learns to play games like chess or Go by playing against itself and improving based on the results of the games.
- **Autonomous Vehicles**: Reinforcement learning is used for decision-making, such as learning optimal driving strategies through interactions with the environment.

---

5. Self-supervised Learning

Definition: Self-supervised learning is a type of learning where the model generates labels from the input data itself, creating pseudo-labels that help the model learn patterns. This method is often considered a form of unsupervised learning with some supervision embedded in the data.

How it works:

- The model uses parts of the input data to predict other parts of the data, essentially creating its own supervision.
- The self-generated labels serve as a kind of "self-supervision" that guides learning, especially when labeled data is scarce.

Examples:

- **Natural Language Processing (NLP)**: In tasks like language modeling, where a model is trained to predict the next word in a sentence, the sentence itself acts as the source of "labels."

- **Contrastive Learning**: Learning representations of data by contrasting positive and negative samples (e.g., in computer vision, predicting whether two images are similar or not).

Ans- 3(a) **Bar Chart**

- **Definition**: A bar chart uses rectangular bars to represent data. The length of each bar is proportional to the value it represents, making it easy to compare different categories or groups.
- **Key Features**:
    - **Axes**: The x-axis typically represents categories, and the y-axis represents numerical values.
    - **Usage**: Bar charts are ideal for comparing discrete data across different categories or groups.
    - **Examples**:
    - Sales comparison between different products.
    - Population of different countries.
    - **Advantages**:
    - Simple and easy to interpret.
    - Useful for categorical data comparison.

2. Line Chart

- **Definition**: A line chart is used to represent data points connected by a continuous line, making it effective for displaying trends over time.
- **Key Features**:
    - **Axes**: The x-axis usually represents time (e.g., days, months), and the y-axis represents the variable of interest.
    - **Usage**: Line charts are ideal for tracking changes over intervals and showing trends.
    - **Examples**:
    - Stock price changes over time.
    - Temperature variations over the year.
    - **Advantages**:
    - Effective in showing data trends.
    - Good for representing continuous data.

- **3. Pie Chart**

  Definition: A pie chart is a circular chart divided into segments, each representing a proportion of the total. It is used to display relative percentages or parts of a whole.
- **Key Features**:
    - **Segments**: Each slice represents a category, and the size of the slice is proportional to the category's percentage of the total.
    - **Usage**: Pie charts are ideal for showing how a total is divided into various components.
    - **Examples**:
    - Market share of different companies.
    - Distribution of a budget across various departments.
    - **Advantages**:
    - Simple to understand.
    - Effective for showing relative proportions.

- **4. Scatter Plot**

Definition: A scatter plot uses points to represent values of two continuous variables. Each point represents an observation, and its position on the chart shows its value on the x and y axes.

- **Key Features**:
    - **Axes**: The x-axis and y-axis represent continuous variables.
    - **Usage**: Scatter plots are used to identify relationships, correlations, or trends between two variables.
    - **Examples**:
    - Relationship between height and weight of individuals.
  - Correlation between advertising spend and sales growth.

- **Advantages**:
- Ideal for identifying correlations or trends.
- Helps in detecting outliers and patterns in data.

Ans 3(b) Dimensionality Reduction Techniques

Dimensionality reduction is the process of reducing the number of features (variables) in a dataset while retaining as much relevant information as possible. It is often used to improve the performance of machine learning models, reduce overfitting, and make the data easier to visualize and analyze. Below are two popular dimensionality reduction techniques:

---

1. Principal Component Analysis (PCA)

Definition:

Principal Component Analysis (PCA) is a statistical technique that transforms high-dimensional data into a lower-dimensional form while retaining most of the variance (information) from the original data. It does this by finding the principal components (the directions of maximum variance) in the dataset.

How it Works:

- Step 1: Standardize the dataset (if the features have different scales) to give each feature equal weight.
- Step 2: Compute the covariance matrix to understand the relationships between the features.
- Step 3: Perform Eigen decomposition to find the eigenvectors (principal components) and eigenvalues (the amount of variance explained by each component).
- Step 4: Select the top k principal components that capture the most variance in the data. The new dataset is formed by projecting the data onto these selected components.

Key Features:

- Reduces the dataset to a smaller number of principal components.
- Maximizes the variance in the reduced space, ensuring minimal loss of information.
- The components are orthogonal, meaning they are uncorrelated.

Examples:

- Reducing the number of variables in image processing tasks while preserving important features (e.g., reducing the dimensions of image data for facial recognition).
- Analyzing high-dimensional genomic data by reducing the number of features.

Advantages:

- Helps in data visualization (e.g., 2D or 3D scatter plots of high-dimensional data).
- Reduces computational cost and storage requirements by eliminating redundant features.

2. t-Distributed Stochastic Neighbor Embedding (t-SNE)

Definition:

t-Distributed Stochastic Neighbor Embedding (t-SNE) is a non-linear dimensionality reduction technique primarily used for the visualization of high-dimensional datasets in two or three dimensions. Unlike PCA, t-SNE is more focused on preserving the local structure of the data rather than maximizing variance.

How it Works:

- Step 1: t-SNE first computes pairwise similarities between points in the high-dimensional space, typically using a Gaussian distribution.
- Step 2: It then tries to find a lower-dimensional representation (2D or 3D) of the data such that similar points in the high-dimensional space are placed closer together in the lower-dimensional space.
- Step 3: t-SNE uses a probability distribution to measure the similarity between points and minimizes the divergence between the original and projected distributions.

Key Features:

- Focuses on preserving the local structure of the data (i.e., it ensures that similar points in high-dimensional space stay close in the lower-dimensional representation).
- Very effective in visualizing clusters or patterns in data.
- Works well with non-linear relationships between features.

Examples:

- Visualizing high-dimensional data in a 2D or 3D plot to reveal hidden clusters, such as in customer segmentation or visualizing complex neural network activations.

- Reducing the dimensions of word vectors in Natural Language Processing (NLP) for visualizing word embeddings.

Advantages:

- Excellent for visualizing complex, high-dimensional data (e.g., in 2D or 3D space).
- Particularly effective when dealing with data that has non-linear relationships.

Disadvantages:

- t-SNE is computationally expensive, especially with large datasets.
- It is primarily a visualization technique and is not suitable for general use in downstream machine learning tasks.

Ans- 4(a)

Step 1: Prepare the Data

We are given two data points:

- $\mathbf{X}_1 = (2, 6)$
- $\mathbf{X}_2 = (3, 7)$

These data points represent two-dimensional data, where each point has two features. We'll treat these as vectors in a 2D space.

Step 2: Standardize the Data

Since the data points are already in a similar range, standardiza
in general, you would standardize if the features have different
with the raw values.

Step 3: Compute the Covariance Matrix

To apply PCA, we first calculate the covariance matrix to unders
each other.

- **Mean of each feature**:
  $\mu_1 = \frac{2+3}{2} = 2.5 \quad \text{(mean of the first feature)}$
  $\mu_2 = \frac{6+7}{2} = 6.5 \quad \text{(mean of the second feature)}$
- **Centered data points** (subtract the mean from each data p
  $\mathbf{X}'_1 = (2 - 2.5, 6 - 6.5) = (-0.5, -0.5)$
  $\mathbf{X}'_2 = (3 - 2.5, 7 - 6.5) = (0.5, 0.5)$

- **Covariance matrix:**

  The covariance matrix is calculated as:

$$\mathrm{Cov}(X) = \frac{1}{n}\sum(\mathbf{X}_i - \mu)(\mathbf{X}_i - \mu)^T$$

  Where n is the number of data points (in this case, 2).

  The covariance matrix becomes:

$$\mathrm{Cov}(X) = \frac{1}{2}\begin{bmatrix}(-0.5) & (0.5)\end{bmatrix}\begin{bmatrix}-0.5 & 0.5\end{bmatrix} = \frac{1}{2}\begin{bmatrix}0.25 & 0.25 \\ 0.25 & 0.25\end{bmatrix}$$

$$\mathrm{Cov}(X) = \begin{bmatrix}0.25 & 0.25 \\ 0.25 & 0.25\end{bmatrix}$$

Step 4: Eigenvalues and Eigenvectors

Next, we need to find the **eigenvalues** and **eigenvectors** of the covariance matrix. This will allow us to determine the principal components.

The eigenvalues λ are solutions to the characteristic equation:

$$\det(\mathrm{Cov}(X) - \lambda I) = 0$$

Where I is the identity matrix.

The covariance matrix is:

$$\mathrm{Cov}(X) = \begin{bmatrix}0.25 & 0.25 \\ 0.25 & 0.25\end{bmatrix}$$

To find the eigenvalues, we compute:

$$\det\left(\begin{bmatrix}0.25 & 0.25 \\ 0.25 & 0.25\end{bmatrix} - \lambda\begin{bmatrix}1 & 0 \\ 0 & 1\end{bmatrix}\right) = \det\left(\begin{bmatrix}0.25-\lambda & 0.25 \\ 0.25 & 0.25-\lambda\end{bmatrix}\right) = 0$$

The characteristic polynomial is:

$$(0.25-\lambda)^2 - 0.25 \times 0.25 = 0$$

$$\lambda^2 - 0.5\lambda = 0$$

$$\lambda(\lambda - 0.5) = 0$$

Thus, the eigenvalues are:

$$\lambda_1 = 0.5 \quad \lambda_2 = 0$$

The corresponding **eigenvectors** can be found by substituting each eigenvalue back into the equation:

$$\text{Cov}(X)\mathbf{v} = \lambda\mathbf{v}$$

For $\lambda_1 = 0.5$, solving the system of equations gives an eigenvector:

$$\mathbf{v_1} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

For $\lambda_2 = 0$, solving the system gives an eigenvector:

$$\mathbf{v_2} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

Step 5: Project the Data onto the Principal Components

The principal component directions are given by the eigenvectors. We project the original data onto these components.

1. The first principal component (corresponding to $\lambda_1 = 0.5$) is the direction of maximum variance, given by $\mathbf{v_1} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.
2. The second principal component (corresponding to $\lambda_2 = 0$) is the direction with zero variance, given by $\mathbf{v_2} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$.

The data points are projected onto the first principal component:

$$\mathbf{X}'_1 \cdot \mathbf{v_1} = (-0.5, -0.5) \cdot (1, 1) = -1$$

$$\mathbf{X}'_2 \cdot \mathbf{v_1} = (0.5, 0.5) \cdot (1, 1) = 1$$

Thus, the transformed data points along the first principal component are:

Ans- 5(a)

Hypothesis testing is a statistical method used to make inferences or draw conclusions about a population based on sample data. It involves testing an assumption (hypothesis) about a population parameter using sample data to determine whether there is enough evidence to support or reject that hypothesis.

Key Concepts in Hypothesis Testing

1. **Null Hypothesis (H0H_0H0)**:
   This is the hypothesis that there is no effect, relationship, or difference in the population. It represents the status quo or a default assumption.
2. **Alternative Hypothesis (HaH_aHa or H1H_1H1)**:
   This hypothesis suggests that there is an effect, relationship, or difference in the population. It is the opposite of the null hypothesis.
3. **Test Statistic**:
   A value calculated from the sample data that is used to decide whether to reject the null hypothesis. Common test statistics include the t-statistic, z-statistic, and chi-square statistic.
4. **P-value**:
   The probability of observing the test statistic or something more extreme, assuming the null hypothesis is true. A smaller p-value indicates stronger evidence against the null hypothesis.
5. **Significance Level (α\alphaα)**:
   The threshold probability at which you reject the null hypothesis. Common values are α=0.05\alpha = 0.05α=0.05,

α=0.01\alpha = 0.01α=0.01, or α=0.10\alpha = 0.10α=0.10.

6. **Decision**:
   - If the **p-value** is less than or equal to the significance level (α\alphaα), you reject the null hypothesis (H0H_0H0).
   - If the **p-value** is greater than α\alphaα, you fail to reject the null hypothesis.
7. **Type I Error (False Positive)**:
   Rejecting H0H_0H0 when it is actually true.
8. **Type II Error (False Negative)**:
   Failing to reject H0H_0H0 when it is actually false.

Ans- 6(a) t-Test

A **t-test** is a statistical test used to determine if there is a significant difference between the means of two groups or between the mean of a single group and a known value. It is particularly useful when the sample size is small or the population standard deviation is unknown.

There are different types of t-tests, such as:

- **One-sample t-test**: Used to compare the mean of a single sample to a known value (e.g., a population mean).
- **Independent two-sample t-test**: Used to compare the means of two independent groups.
- **Paired sample t-test**: Used to compare the means of two related groups (e.g., before and after treatment).

Steps to Perform a t-Test:

1. **State the Hypotheses**:
   - Null Hypothesis (H0H_0H0): There is no significant difference between the means.
   - Alternative Hypothesis (HaH_aHa): There is a significant difference between the means.
2. **Choose the Significance Level (α\alphaα)**:
   Typically, α=0.05\alpha = 0.05α=0.05 (5%).
3. **Compute the Test Statistic (t-statistic)**: The formula for the t-statistic depends on the type of t-test used:
   - For a one-sample t-test: t=x¯−μs/nt = \frac{\bar{x} - \mu}{s / \sqrt{n}}t=s/nx¯−μ Where:
     - x¯\bar{x}x¯ = sample mean
     - μ\muμ = population mean (or known value)
     - sss = sample standard deviation
     - nnn = sample size
4. **Determine the Degrees of Freedom**:
   Degrees of freedom (dfdfdf) are typically n−1n - 1n−1 for a one-sample t-test, where nnn is the sample size.
5. **Find the Critical Value**:
   The critical value is determined from the t-distribution table, based on the chosen significance level (α\alphaα) and degrees of freedom.
6. **Compare the Test Statistic with the Critical Value**:
   - If the absolute value of the test statistic is greater than the critical value, reject the null hypothesis.
   - If the test statistic is less than the critical value, fail to reject the null hypothesis.
7. **Interpret the Results**:
   Based on the comparison, conclude whether the difference between the sample mean and the population mean is statistically significant.

Example of a t-Test: One-Sample t-Test

Scenario:

A factory claims that the average weight of a batch of apples is 150 grams. A researcher collects a sample of 10 apples and finds that the sample mean weight is 145 grams with a sample standard deviation of 8 grams. Is the sample mean significantly different from 150 grams?

Step 1: State the Hypotheses

- Null Hypothesis (H0H_0H0): The average weight of apples is 150 grams (μ=150\mu = 150μ=150).
- Alternative Hypothesis (HaH_aHa): The average weight of apples is not 150 grams (μ≠150\mu \neq 150μ□=150).

Step 2: Choose the Significance Level

- Choose α=0.05\alpha = 0.05α=0.05.

Step 3: Compute the t-statistic Using the formula:

Step 4: Determine the Degrees of Freedom Degrees of freedom for a one-sample t-test is n−1n - 1n−1, where n=10n = 10n=10:

Step 5: Find the Critical Value Using a t-distribution table or calculator for α=0.05\alpha = 0.05α=0.05 and df=9df = 9df=9,

Step 6: Compare the Test Statistic with the Critical Value

- The calculated t-statistic is −1.98-1.98−1.98.

- The critical value is ±2.262\pm 2.262±2.262.

Since −1.98-1.98−1.98 is not greater than 2.2622.2622.262 (in absolute value), we **fail to reject the null hypothesis**.
Step 7: Conclusion There is not enough evidence to reject the null hypothesis. The sample mean of 145 grams is not significantly different from the population mean of 150 grams at the 0.05 significance level.
Ans- 6(b) **Supervised vs. Unsupervised Machine Learning**
Machine learning can be broadly categorized into two types based on the nature of the data and the learning process: **supervised learning** and **unsupervised learning**. Here's a comparison of both:

---

1. Supervised Learning
In supervised learning, the model is trained on labeled data. This means that the training dataset includes both input data and the corresponding correct output (also called labels). The goal is for the algorithm to learn the mapping from inputs to outputs so that it can make predictions on unseen data.
Key Characteristics of Supervised Learning:
- **Labeled Data**: The training data contains input-output pairs (i.e., each input is associated with a known label or target value).
- **Goal**: To learn a mapping from inputs to outputs. In regression tasks, the model predicts a continuous value. In classification tasks, the model predicts discrete labels.
- **Learning Process**: The model uses the known labels to adjust its parameters and minimize the difference between predicted and actual outcomes (e.g., using a loss function).
- **Example Tasks**:
  - **Classification**: Categorizing emails as spam or not spam.
  - **Regression**: Predicting house prices based on features like size and location.

Examples of Supervised Learning Algorithms:
- **Linear Regression**
- **Logistic Regression**
- **Support Vector Machines (SVM)**
- **Decision Trees**
- **Random Forest**
- **k-Nearest Neighbors (k-NN)**

Pros:
- The presence of labeled data allows the model to learn with clear guidance.
- Easier to evaluate and tune models since performance can be directly compared to known labels.

Cons:
- Requires a large amount of labeled data, which can be expensive or time-consuming to obtain.
- The model may not generalize well to unseen data if it overfits the training data.

---

2. Unsupervised Learning
In unsupervised learning, the model is trained on **unlabeled data**, meaning there are no predefined labels or target values. The goal is for the algorithm to find hidden patterns or structures in the input data without explicit guidance.
Key Characteristics of Unsupervised Learning:
- **Unlabeled Data**: The training data consists of inputs only, with no corresponding labels or outcomes.
- **Goal**: To uncover underlying structures, patterns, or groupings in the data (e.g., clustering or dimensionality reduction).
- **Learning Process**: The model tries to learn the intrinsic structure of the data and group similar inputs together or reduce the number of features.
- **Example Tasks**:
  - **Clustering**: Grouping customers based on purchasing behavior.
  - **Dimensionality Reduction**: Reducing the number of features while preserving important patterns (e.g., PCA).

Examples of Unsupervised Learning Algorithms:
- **k-Means Clustering**
- **Hierarchical Clustering**
- **Principal Component Analysis (PCA)**
- **Autoencoders**
- **Gaussian Mixture Models (GMM)**

Pros:
- No need for labeled data, making it easier and cheaper to work with real-world datasets.
- Useful for exploratory data analysis and uncovering hidden patterns or relationships.

Cons:

- Harder to evaluate the model's performance since there's no ground truth.
- The results might be less interpretable compared to supervised learning tasks.

Comparison Table

Aspect	**Supervised Learning**	**Unsupervised Learning**
Data	Labeled data (input-output pairs)	Unlabeled data (only input data)
Goal	Learn a mapping from inputs to outputs	Discover hidden patterns or structures in data
Learning Type	Prediction-based (classification or regression)	Pattern-based (clustering, dimensionality reduction)
Output	Predicted labels (for classification) or values (for regression)	Groups or reduced representations of data
Evaluation	Can be directly evaluated using performance metrics like accuracy or RMSE	Harder to evaluate since there's no ground truth
Examples	Email spam detection, stock price prediction	Customer segmentation, anomaly detection
Algorithms	Linear Regression, SVM, k-NN, Random Forest, etc.	k-Means, PCA, DBSCAN, Autoencoders, etc.
Data Requirement	Requires labeled data (expensive or time-consuming)	Does not require labeled data (easier data collection)